iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59537

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Visual Question Generation from Remote Sensing Images Using Gemini API

M. Kamala[1], B. Pravalika[2], Y. Laxmi Narayana[3], P. Arya Patel[4]
UG Student, Department of Computer Science & Engineering, CMR College of Engineering & Technology, Hyderabad, India

Abstract: ***Visual Question Generation Extracting Information from Remote Sensing Images Remote Sensing Images plays a vital role in understanding and extracting information from aerial and satellite images. Utilizing Bidirectional Encoder Representation from Transformers (BERT) for extracting valuable insights from remote sensing images. Gemini Application Programming Interface(API), and Convolution Neural Networks (CNNs) are used. First, The proposed methodology employs CNN to extract high-level features from remote sensing images, capturing spatial data and generating questions. Similarly, the Gemini Application Programming Interface(API) integrates contextual understanding into the question-generation process by providing relevant environmental data. Lastly, BERT functions as a language model in which employees enhance and refine the generated questions by taking into account both the syntax and semantics. Hence, by combining all these techniques we are capable of generating required relevant questions from remote sensing images in an enhanced and efficient way.***

Index Terms: Visual Question Generation, CNN, Gemini API, Remote Sensing Images, Natural Language Processing, Deep Learning, BERT.

I. INTRODUCTION

The burgeoning field of the Visual Address Era (VQG) points to bridging the crevice between visual question substance and normal dialect by naturally producing questions almost pictures. This paper proposes a novel approach joining Convolutional Neural Systems (CNNs), the Gemini API, and Utilizing Bidirectional Encoder Representation from Transformers for creating questions from further detecting pictures. CNNs investigate visual highlights, Gemini API enhances relevant understanding, and BERT refines semantics. This strategy upgrades address pertinence and specificity, engaging clients to extricate experiences from endless visual information stores.

Xiong, Z., Zhang, F., Wang, Y., Shi, Y., & Zhu, X. X [1] Soil perception, pointing at observing the condition of planet Soil utilizing further detecting information, is basic for progressing our everyday lives and living environment. With a developing number of satellites in a circle, an expanding number of datasets with assorted sensors and inquiries about spaces are being distributed to encourage the exploration of the inaccessible detecting community Huang, W., Wang, Q., & Li, X [2] With the benefits of profound learning innovation, creating captions for inaccessible detecting pictures has gotten to be achievable, and awesome advances have been made in this field over the later long time. Be that as it may, a large-scale variety of further detecting pictures, which would lead to mistakes or exclusions in including extraction, still limits the assist advancement of caption quality. Fu, K., Li, Y., Zhang, W., Yu, H., & Sun, X [3] The encoder–decoder system has been broadly utilized within the further detecting picture captioning task. When we have to extricate inaccessible detecting pictures containing specific characteristics from the portrayed sentences for investigation, wealthy sentences can make strides the ultimate extraction comes about. Be that as it may, the Long Short-Term Memory (LSTM) organize utilized in decoders still loses a few data within the picture over time when the produced caption is long instrument is predominant in this errand but still has a few disadvantages. The customary consideration instrument as it were employments visual data that is almost inaccessible in detecting pictures.

Zhang, Z., Diao, W., Zhang, W., Yan, M., Gao, X., & Sun, X [4] Critical advances have been made in further detecting picture captioning by encoder-decoder systems. The customary consideration without considering utilizing the name data to direct the calculation of consideration covers.

II. LITERATURE SURVEY

Shi, Z., & Zou, Z. [5] This paper examines a charming address within the inaccessible detecting field: “Can a machine produce humanlike dialect portrayals for a further detecting image The modified delineation of a farther-detecting picture (to be particular, blocked off recognizing picture captioning) may be a basic but rarely considered errand for fabricated bits of knowledge? It is more challenging as the depiction must not as it were capture the ground components of diverse scales, but moreover express their traits as well as how these components connected.

Zhang, Z., Diao, W., Zhang, W., Yan, M., Gao, X., & Sun, X [6] Critical advance has been made in inaccessible detecting picture captioning by encoder-decoder systems. The customary consideration instrument is predominant in this errand but still has a few downsides. The customary consideration component as it were employments visual data almost the further detecting pictures without considering utilizing the name data to direct the calculation of consideration covers. To this conclusion, a novel consideration component, specifically Label-Attention Instrument (LAM), is proposed in this paper. LAM moreover utilizes the name data of high-resolution inaccessible detecting pictures to create characteristic sentences to depict the given pictures.

Uppal, S., Madan, A., Bhagat, S., Yu, Y., & Shah, R. R [7] Visual Address Era (VQG) is the errand of creating common questions based on a picture. Fan, Z., Wei, Z., Li, P., Lan, Y., & Huang, X. [8] In our system, an address is developed in two steps. To begin with, an address sort is examined to decide what kind of data is asked. Moment, the substance of the address is produced conditioning on the tested address sort and the visual data of the picture. Prevalent strategies in the past have to investigate image-to-sequence designs prepared with the most extreme probability which have illustrated significant produced questions given a picture and its related ground-truth reply. VQG gets to be more challenging if the picture contains wealthy relevant data portraying its distinctive semantic categories

Abdullah, T., Bazi, Y., Al Rahhal, M. M., Mohali, M. L., Rangarajan, L., & Zuhair, M. [9] The unfaltering availability of inaccessible detecting information, especially tall determination pictures, has vivified exceptional inquiries about yields within the further sensing community. Two of the foremost dynamic points in this respect allude to picture classification and recovery [1,2,3,4,5]. Picture classification points to relegate scene pictures to a discrete set of arrival use/land cover classes depending on the picture substance [6,7,8,9,10]. As of late, with quickly extended further detecting innovations, both the amount and quality offurther detecting information have been expanded.

III. METHODOLOGY

A. Technologies used

Here, we have used various technological stuff usage of a pre-trained model Gemini API, CNN, and Bert to generate accurate questions from remote sensing images.

B. Gemini API

Gemini offers to get to a run of expansive dialect models, each with its qualities in the address era. You'll be able to select the LLM that best suits your particular needs, like producing open-ended questions, genuine requests, or imaginative prompts. The API acknowledges different input designs, counting content sections, pictures, code bits, or indeed conversational transcripts. This permits you to create questions based on diverse sorts of data, making it flexible for different applications. The API consistently coordinates with other instruments and stages, permitting you to consolidate address eras into different workflows.

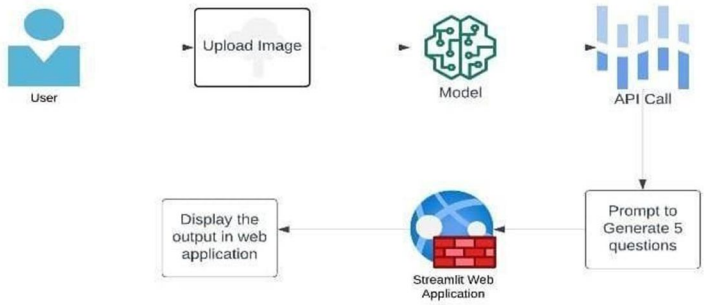


Fig 1: Proposed Methodology Architecture

C. BERT

1) *Gets the Picture Information*: After Gemini API bridges the hole, BERT takes the reins, accepting the complicated pointsof interest extricated from the inaccessible detecting picture.
2) *Interprets Visual Designs:* Like a prepared criminologist, BERT analyzes the picture, leveraging its tremendous informationon visual components and their connections. This incorporates arrival cover sorts, surfaces, shapes, and spatial courses of action.
3) *Makes Curious Questions:* Based on its investigation, BERT changes the visual information into characteristic dialect questions. These questions act as prompts, welcoming a more profound investigation of the image's substance and potential suggestions.

4) *Convolutional Neural Networks:* It Organizes, another sort of neural arrangement design. CNNs exceed expectations at recognizing designs in pictures, making them well-suited for assignments like picture classification and protest location. In this case, CNN likely analyzes the transferred picture to extricate visual highlights that advise the address era preparation.
5) *Input Picture*: The client transfers a picture to the framework.
6) *Gemini Vision Professional API Call:* The picture is sent to the Gemini Vision Professional API. The API analyzes the picture and produces five questions almost it.
7) *Streamlit Integration:* The created questions are shown to the client in a web application.
8) *Show the yield in the web application:* The point-by-point depiction of the picture is at that point shown to the client within the web application.

The Gemini API engages designers with a vigorous set of functionalities to consistently coordinate cryptocurrency exchange and account administration into their applications. Advertising both REST and WebSocket APIs, clients pick up get to real-time showcase information, arrange arrangements, and account administration highlights. With secure verification through API keys, engineers can certainly associate with the trade, guaranteeing information judgment and client security. Gemini's comprehensive documentation and back framework advance improve the improvement involvement, empowering the creation of modern exchanging calculations, portfolio administration apparatuses, and showcase examination stages.

D. Implementation of Block Diagram

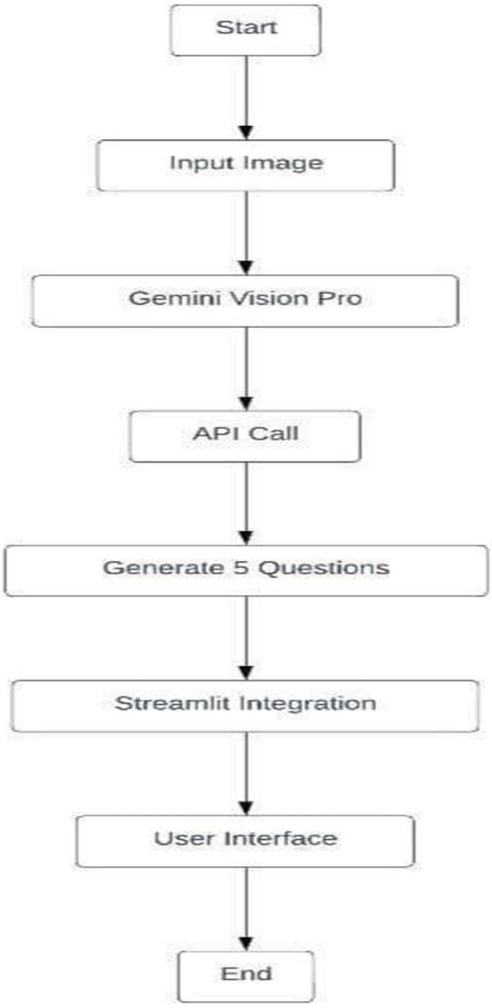


Fig 2: Execution Flow of Proposed Solution

IV. RESULTS AND DISCUSSION

A. Figures

Initially, when we run the required commands the user interface Is opened.

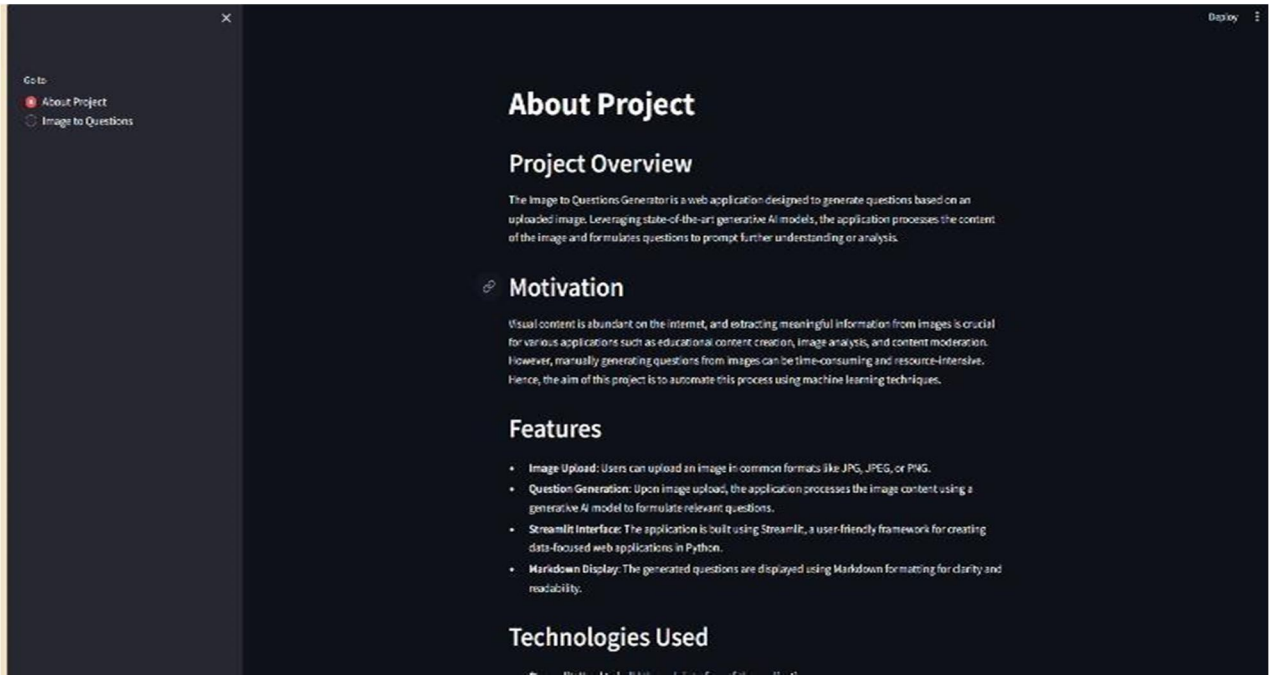


Fig 3: User Interface

Once the interface is opened then we need to upload an image in the provided user interface.

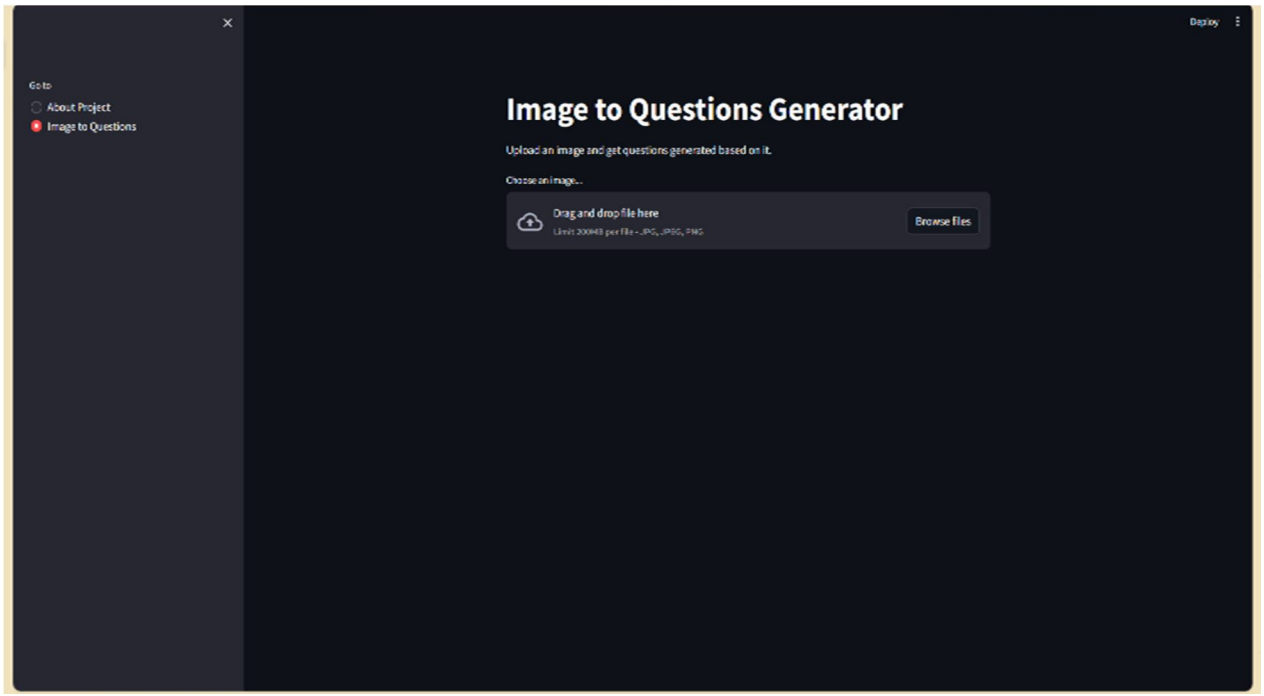


Fig 4: Uploading Of Image

Later on, using pre-trained models the features required are extracted from the image uploaded and then the required output is displayed on the screen

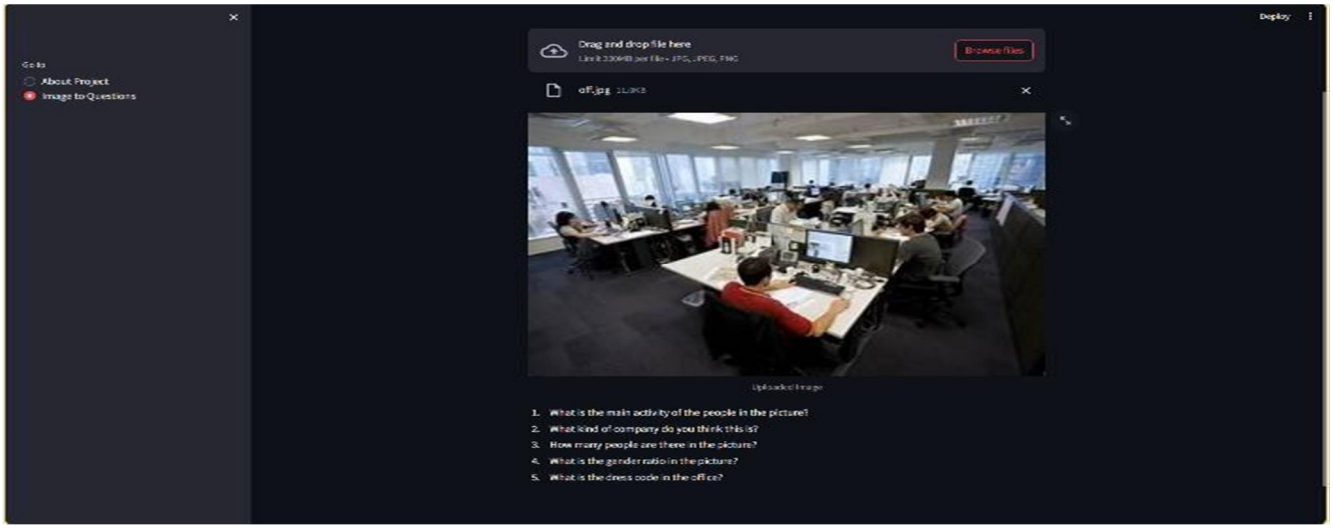


Fig 5: Extract Features and Generate Questions (eg1).

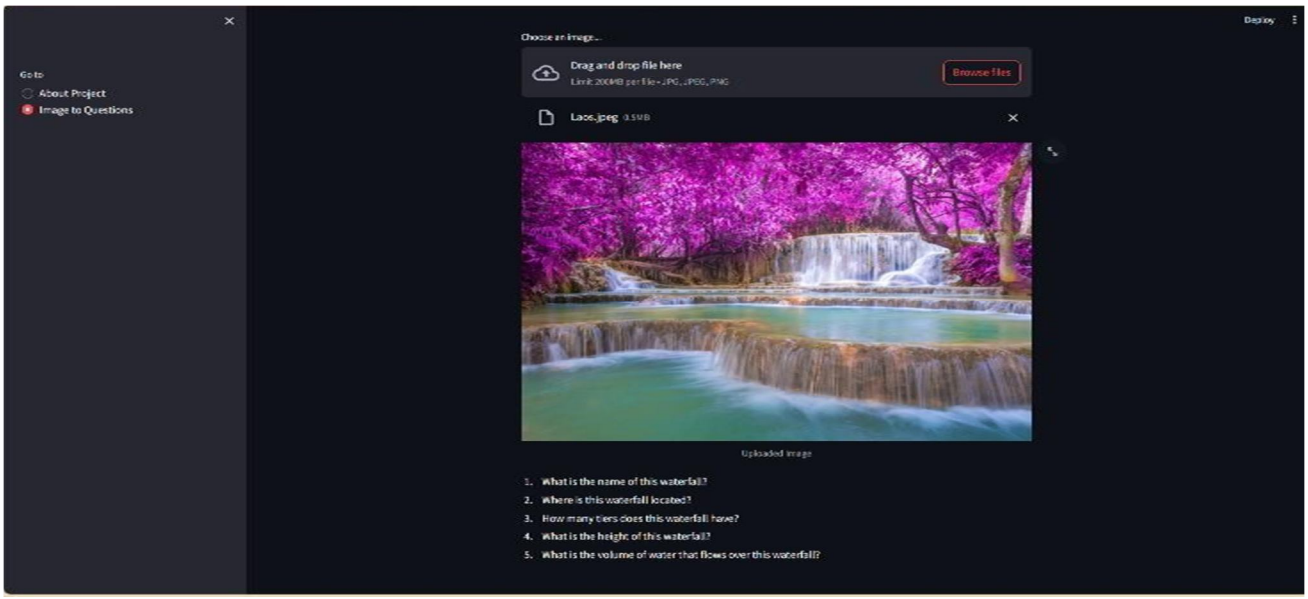


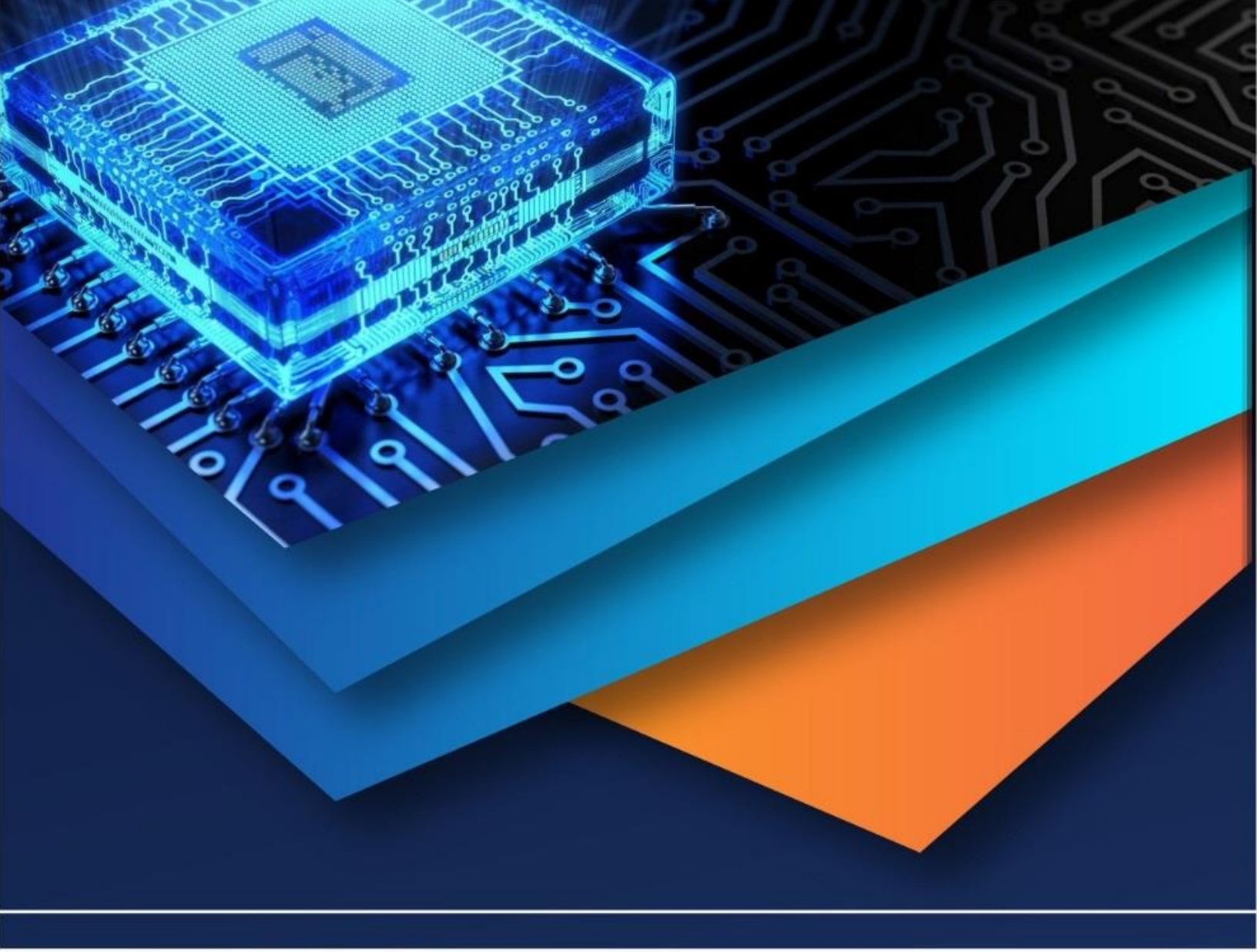
Fig 6: Extract Features and Generate Questions(eg2).

V. CONCLUSION

Visual code time is a new era of visual interaction that can leverage advances in machine learning in unprecedented ways. For example, it can be used to detect complexity in images making it possible to draw queries directly from the visual features. This approach simplifies speech creation and incorporates advanced techniques in computer vision and natural language processing to provide advanced interactivity in machine learning algorithms by extracting visual cues and by the extraction of meaningful information. Potential impacts help businesses in market research from educational seminars to image-based questionnaires or content analysis tools among others including how interesting cross-industry collaborations and partnerships (partnerships) have developed development through the integration of computer vision and natural language processing The development further highlights the potential of machine learning to modify our perception, interpretation, and communication of visual code time as presented here.

REFERENCES

[1] Xiong, Z., Zhang, F., Wang, Y., Shi, Y., & Zhu, X. X. (2022). Earth nets: Empowering AI in earth observation. arid preprintarXiv:2210.04936.
[2] Huang, W., Wang, Q., & Li, X. (2020). Denoising-based multiscale feature fusion for remote sensing image captioning.Geoscience and Remote Sensing Letters, 18(3), 436-440.
[3] Fu, K., Li, Y., Zhang, W., Yu, H., & Sun, X. (2020). Boosting memory with a persistent memory mechanism for remote sensing image captioning. Remote Sensing, 12(11), 1874.
[4] Zhang, Z., Diao, W., Zhang, W., Yan, M., Gao, X., & Sun, X. (2019). LAM: Remote sensing image captioning with the label-attention mechanism. Remote Sensing, 11(20), 2349.
[5] Shi, Z., & Zou, Z. (2017). Can a machine generate humanlike language descriptions for a remote-sensing image? IEEETransactions on Geoscience and Remote Sensing, -55(6), 3623-3634.
[6] Zhang, Z., Diao, W., Zhang, W., Yan, M., Gao, X., & Sun, X. (2019). LAM: Remote sensing image captioning with thelabel-attention mechanism. Remote Sensing, 11(20), 2349.
[7] Uppal, S., Madan, A., Bhagat, S., Yu, Y., & Shah, R. R. (2021, March). C3VQG: Category consistent cyclic visual questiongeneration. In Proceedings of the 2nd ACM International Conference on Multimedia in Asia (pp. 1-7).
[8] Fan, Z., Wei, Z., Li, P., Lan, Y., & Huang, X. (2018, July).
[9] A Question Type Driven Framework to Diversify Visual Question Generation. In IJCAI (pp. 4048-4054).
[10] Abdullah, T., Bazi, Y., Al Rahhal, M. M., Mohali, M. L., Rangarajan, L., & Zuhair, M. (2020). Texters: Deepbidirectional triplet network for matching text to remote sensing images Remote Sensing, 12(3), 405.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY